93-28

WORKING NOTES

AAAI
FALL SYMPOSIUM SERIES

Symposium:
Machine Learning in Computer Vision: What, Why, and How?

Program Committee:
Kevin Bowyer, University of South Florida, Cochair
Lawrence Hall, University of South Florida, Cochair
Chris Brown, University of Rochester
Bruce Draper, University of Massachusetts
Tom Mitchell, Carnegie-Mellon University
Dean Pomerleau, Carnegie-Mellon University
Larry Rendell, University of Illinois

OCTOBER 22, 23, 24, 1993

SHERATON IMPERIAL HOTEL & CONVENTION
CENTER, RALEIGH, NORTH CAROLINA

1993 Fall Symposium Series

Machine Learning in Computer Vision: What, Why, and How?

Working Notes

(Distribution limited to symposium attendees. Not for citation.)

October 22 - 24, 1993

Sheraton Imperial Hotel & Convention Center

Raleigh, North Carolina

Sponsored by the

American Association for Artificial Intelligence

Table of Contents

ISSUES IN LEARNING FROM NOISY SENSORY DATA

Jerzy W. Bala* and Peter W. Pachowicz†

*† Center for Artificial Intelligence
† System Engineering Department

George Mason University
Fairfax, VA 22030

{bala, ppach}@aic.gmu.edu

Abstract

This paper discusses issues in noise tolerant learning from sensory data. A model driven approach to symbolic learning from noisy data is suggested.

Introduction

Sensor-driven characteristics of visual objects are rarely noise free and most often quite noisy. "*The visual world is noisy. Even well posed visual computations are often numerically unstable, if noise is present in both the scene and the image. Scenes are usually corrupted by "noise" coming from various sources (dust, fog, sun glitter, etc.). The image formation process introduces additional noise. As a result, many problems which theoretically have unique solutions become very unstable in the presence of input noise.*" [Rosenfeld, Aloimonos et al., 1990]. To explore the application of learning techniques to vision domains, it is, therefore, very important to develop approaches that are successful despite a high level of noise in the data.

As indicated by Quinlan [Quinlan, 1986], regardless of the source, noise can be expected to affect the formation and use of classification rules. Inductive learning systems must perform some form of generalization in order to anticipate unseen examples. Ideally, a concept description generated by an inductive learning system should cover all examples (including unseen examples) of the concept (completeness) and no examples of other concepts (consistency), and this is what most inductive learning systems do, generating a complete and consistent concept description which covers all positive examples and no negative examples. In the case of noisy data, complete and consistent descriptions are problematic because multiple concept descriptions can partially overlap in the attribute space. This is so, because attribute noise skews the distribution of attribute value from the correct value. Because of the existence of noise in the sensory data, some positive examples are noise, that is, they are actually negative examples. We call such examples "positive noisy examples". These positive noisy examples are covered by the complete and consistent description. These examples, however, should not be covered. On the other hand, some negative examples can be noise, i.e., they are actually positive examples. Such negative examples are referred to as "negative noisy examples". Negative noisy examples are incorrectly left uncovered.

There are two basic groups of approaches to learning from noisy data. One is to allow a certain degree of inconsistent classification of training examples so that the descriptions will be general enough to describe basic characteristics of a concept. This approach has been taken by the ID family of algorithms [Quinlan, 1986]. The main noise-handling mechanism for decision trees is tree pruning. There are two types of tree pruning: pre-pruning, performed during the construction of a decision tree, and post-pruning, used after the decision tree is constructed. The second approach is to discard some of the unimportant rules and retain those covering the largest number of examples. The remaining rules are a general description of the concept. Typical algorithms using these techniques are the AQ family of algorithms [Michalski, 1986]. Rule truncation in AQ15 ([Michalski, 1986, Zhang and Michalski, 1989, Pachowicz and Bala, 1991], [Michalski, 1986, [Zhang and Michalski, 1989], and the significance test in CN2 [Clark and Niblett, 1989] are also examples of that approach. Other approaches are based on the minimum description length principle [Quinlan, 1989] and cross validation to control over-fitting during a training phase [Breiman, Friedman et al., 1984].

Because the above methods try to remove noise in one step, they share a common problem - the final descriptions are based on the initial noisy data. For methods applying pre-pruning, the attributes used to split the instance space in a pruned tree are selected from initial noisy training data. For methods applying pre-truncation, the search for the best disjunct is also influenced by noise in training data. This problem is more severe for post-pruning and post-truncation. The post learning concept optimization cannot merge broken (by noisy examples) concept components (i.e., disjuncts, subtrees). So, the "gaps" in concept descriptions remain unfilled. Post learning optimization will cause the complexity of concept descriptions to decrease only if the concept components are eliminated; i.e., without reorganizing concept descriptions.

Method

A new approach to noise-tolerant learning, called model-driven learning, is presented. The approach was developed to acquire concept descriptions of visual objects from the teacher- or system-provided noisy training samples. This method has the following steps (Figure 1): (i) learning concept descriptions, (ii) the evaluation of learned class descriptions and the detection of less significant disjuncts/subtrees which do not likely represent patterns in the training data, (iii) the iterative removal (e.g., truncation or pruning) of detected disjuncts/subtrees, and (iv) the filtration of training data by optimized rules/trees (i.e., removal of all examples not covered by truncated of pruned concept description). The first novel aspect of this approach is that rules/trees optimized through disjunct/subtree removal are used to filter noisy examples, and then the filtered set of training data is used to relearn improved rules. The second novel aspect is that noise detection is done on the higher level (model level) and can be more effective than traditional data filtration applied on the input level only. The expected effect of such a learning approach is the improvement of recognition performance and a decrease in the complexity of learned class descriptions.

Experimental validation

The experimental investigation and validation of this approach was tested in the domain of texture recognition. The texture domain is always characterized by a high degree of attribute noise and local overlap of class characteristics. Image sections of 12 texture classes were used in experiments. Different texture areas were provided during the learning and recognition phases. Texture attributes were extracted using Laws' method which analyses a local subset of pixels. This subset is defined by a small window in order to derive local characteristics of pixels called texture energy. 300 examples were selected randomly for each texture class. This set was divided into three separate data sets, where each set consisted of 200 training examples and 100 testing examples (1 out of 3 data selection method). Learning and recognition processes were then run on these three sets separately.

An example of one of the experiment is shown in Figure 2. Experiment was performed using two learning algorithms - the C4.5 decision tree learning program and the AQ16 rule learning program. In the experiment, one iteration of the algorithm was used (initial learning and one relearning from new filtered training data).

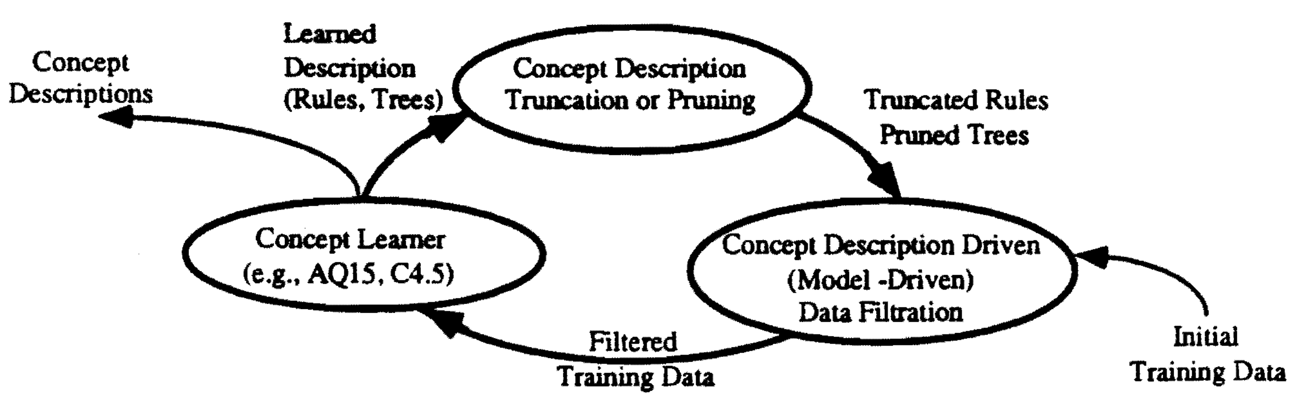

Figure 1: A model-driven noise tolerant learning

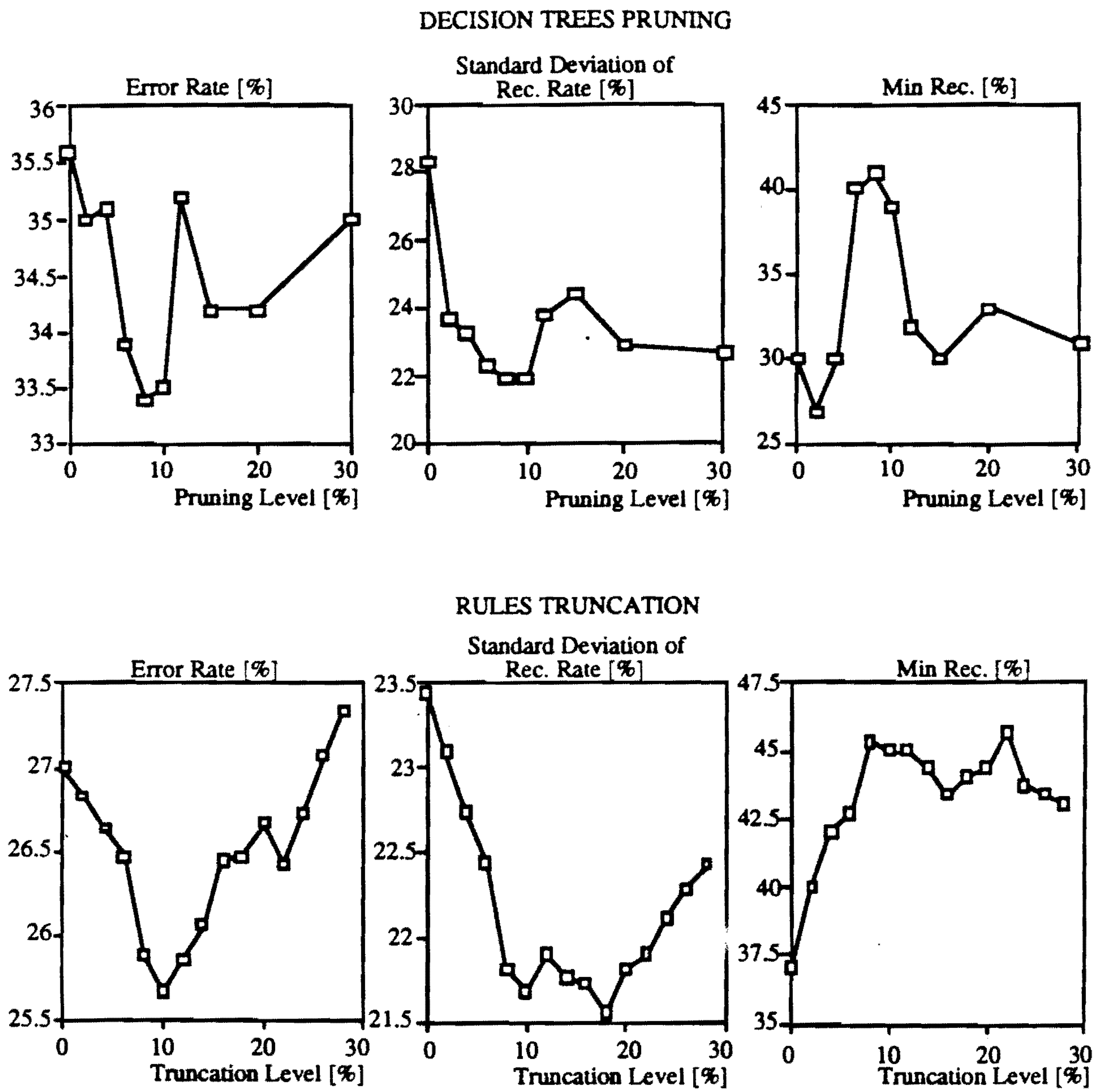

Figure 2: An example of the results for one iteration of the method

Different levels of truncation and pruning (Figure 2) correspond to different levels of concept description component removal. The diagrams illustrate the following evaluation criteria: average error rate (the rate of incorrectly classified examples to the total number of testing examples, standard deviation from the average recognition rate, and minimum recognition rate (indicating the worst performing class). An improvement of all evaluation criteria is observed for the 1% to 8% range of pruning level and the 1% to 10% range of truncation. The results obtained for optimal pruning and truncation level outperformed the results obtained from the non-iterative approach.

Acknowledgement

This research was supported in part by the Defense Advanced Research Projects Agency under the grant No. F49620-92-J-0549, administered by the Air Force Office of Scientific Research, and the grant No. N00014-91-J-1854, administrated by the Office of Naval Research, in part by the Office of Naval Research under the grant No. N00014-91-J-1351, and in part by the National Science Foundation under the grant No. IRI-9020266.

References

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, 1984.

Clark, P. and T. Niblett, "The CN2 Induction Algorithm," *Machine Learning*, Vol. **3**, pp. 261-284, 1989.

Michalski, R. S., Mozetic, I., Hong, J., and Lavrac, N., "The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains," *Proceedings of AAAI-86*, Philadelphia, PA, 1986.

Pachowicz, P. and J. Bala, "Texture Recognition Through Machine Learning and Concept Optimization," P91-12, MLI 6, Center for Artificial Intelligence, George Mason University, 1991.

Quinlan, J. R., and Rivest, R.L., "Inferring Decision Trees Using the Minimum Description Length Principle," *Information and Computation*, Vol. **80**, no. 3, 1989.

Quinlan, J. R., *The Effect of Noise on Concept Learning*, Machine Learning: An Artificial Intelligence Approach, Vol. II, R.S. Michalski, J.G. Carbonell and T.M. Mitchell (Eds.), Morgan Kaufmann, Los Altos, CA, 1986.

Rosenfeld, A., J. Aloimonos and L. Davis, "Maryland Progress in Image Understanding," *DARPA Image Understanding Workshop*, 1990.

Zhang, J. and R. S. Michalski, "A Description of Preference Criterion in Constructive Learning: A Discussion of Basic Issues," *Proceedings of the 6th International Workshop on Machine Learning*, Segre, A., Cornell University, Ithaca, NY, 1989.

Zhang, J. and R. S. Michalski, "Rule Optimization via SG-TRUNC Method," *Proceedings of EWSL-89*, Montpellier, France, 1989.